# Parallel Bit Interleaved Coded Modulation

Amir Ingber† and Meir Feder
Department of EE-Systems, Tel Aviv University
Tel Aviv 69978, ISRAEL
{ingber, meir}@eng.tau.ac.il



**Abstract**

A new variant of bit interleaved coded modulation (BICM) is proposed. In the new scheme, called *parallel BICM*, $L$ identical binary codes are used in parallel using a mapper, a newly proposed finite-length interleaver and a binary dither signal. As opposed to previous approaches, the scheme does not rely on any assumptions of an ideal, infinite-length interleaver. Over a memoryless channel, the new scheme is proven to be equivalent to a *binary* memoryless channel. Therefore the scheme enables one to easily design coded modulation schemes using a simple binary code that was designed for that binary channel. The overall performance of the coded modulation scheme is analytically evaluated based on the performance of the binary code over the binary channel. The new scheme is analyzed from an information theoretic viewpoint, where the capacity, error exponent and channel dispersion are considered. The capacity of the scheme is identical to the BICM capacity. The error exponent of the scheme is numerically compared to a recently proposed mismatched-decoding exponent analysis of BICM.

## I. Introduction

Bit interleaved coded modulation (BICM) is a pragmatic approach for coded modulation [1]. It enables the construction of nonbinary communication schemes from binary codes by using a long bit interleaver that separates the coding and the modulation. BICM has drawn much attention in recent years, because of its efficiency for wireless and fading channels.

The information-theoretic properties of BICM were first studied by Caire et. al. in [2]. BICM was modeled as a binary channel with a random state that is known at the receiver. The state determines how the input bit is mapped to the channel, along with the other bits that are assumed to be random. Under the assumption of an infinite-length, ideal interleaver, the BICM scheme is modeled by parallel uses of independent instances of this binary channel. This model is referred to as the *independent parallel channel model*.

Using this model the capacity of the BICM scheme could be calculated. It was further shown that BICM suffers from a gap from the full channel capacity, and that when Gray mapping is used this gap is generally small. In [2], methods for evaluating the error probability of BICM were proposed, which rely on the properties of the specific binary codes that were used (e.g. Hamming weight of error events).

A basic information-theoretic quantity other than the channel capacity is the error exponent [3], which quantifies the speed at which the error probability decreases to zero with the block length $n$. Another tool for evaluating the performance at finite block length is the channel dispersion, which was presented in 1962 [4] and was given more attention only in recent years [5], [6]. It would therefore be interesting to analyze BICM at finite block length from the information-theoretic viewpoint.

Several attempts have been made to provide error exponent results for BICM.

In their work on multilevel codes, Wachsmann et. al. [7] have considered the random coding error exponent of BICM, by relying on the independent parallel channels model. However, there were several flaws in the derivation:

- The independent parallel channels model is justified by an infinite-length interleaver. Therefore it might be problematic to use its properties for evaluating the finite length performance of the BICM scheme. In the current paper we address this point and propose a scheme with a finite-length interleaver for that purpose.
- There was a technical flaw in the derivation, which resulted in an inaccurate expression for the random coding error exponent. We discuss this point in detail in Theorem 4.
- As noticed in [8], the error exponent result obtained in [7] sometimes may even exceed that of unconstrained coding over the channel (called in [8] the "coded modulation exponent"). We therefore agree with [8] in the claim that "the independent parallel channel model fails to capture the statistics of the channel". However, by properly designing the communication scheme the model can become valid in a rigorous way, as we show in Theorem 1.

In [8] (see also [9]), Martinez et al. have considered the BICM decoder as a mismatched decoder, which has access only to the log-likelihood values (LLR) of each bit, where the LLR calculation assumes that the other bits are random, independent and equiprobable (as in the classical BICM scheme [2]). Using results from mismatched decoding, they presented the generalized error exponent and the generalized mutual information, and pinpointed the loss of BICM that incurs from using the mismatched LLRs. Note that when a binary code of length $n$ is used, the scheme requires only $n/L$ channel uses. While this result is valid for any block size and any interleaver length, achieving this error exponent in practice requires complex code design.

† A. Ingber is supported by the Adams Fellowship Program of the Israel Academy of Sciences and Humanities.

For example, one cannot design a good binary code for a binary memoryless channel and have any guarantee that the BICM scheme will perform well with that code. In fact, the code design for this scheme requires taking into account the memory within the levels, or equivalently, nonbinary codes, which is what we wish to avoid when choosing BICM.

On the theoretical side, another drawback of existing approaches is the lack of converse results (for either capacity or error exponent). The initial discussion of BICM information theory in [2] assumes the model of independent channels, and any converse result based on this model must assume that an infinite, ideal, interleaver. Therefore the converse results (such as upper bound on the achievable rate with BICM) do not hold for finite-length interleavers. The authors in [8] provide no converse results for their model.

In this paper we propose the *parallel BICM* (PBICM) scheme, which has the following properties. First, the scheme includes an explicit, *finite length* interleaver. Second, in order to attain good performance on any memoryless channel, PBICM allows one to design a binary code for a binary memoryless channel, and guarantees good performance on the nonbinary channel. Third, because the scheme does not rely on the use of an infinite-length interleaver, the error exponent and the dispersion of the scheme can be calculated (both achievability and converse results) as means to evaluate the PBICM performance at finite block length.

The comparison between PBICM and the mismatched decoding approach [8] should be done with care. With PBICM, when the binary codeword length is $n$ the scheme requires $n$ channel uses. Therefore when the latency kept equal for both schemes, PBICM uses a codeword length that is $L$ times shorter than the codeword used in the mismatched decoder. A fair comparison would be to fix the binary codeword length $n$ for both schemes, resulting in different latency, but equal decoder complexity.

The results presented in the paper are summarized as follows:

- The PBICM communication framework is presented. Over a memoryless channel, it is shown to be equivalent to a *binary* memoryless channel (Theorem 1).
- In Theorem 2, the capacity of PBICM is shown to be equal to the BICM capacity, as calculated in [2].
- PBICM is analyzed at finite block length. The error exponent of PBICM is defined and bounded by error exponent bounds of the underlying binary channel (Theorems 3 and 4).
- The PBICM dispersion is defined as an alternative measure for finite-length performance. It is calculated by the dispersion of the underlying binary channel (Theorems 5 and 6).
- The error exponent of PBICM is numerically compared to the mismatched-decoding error exponent of BICM [8]. The additive white Gaussian noise (AWGN) channel and the Rayleigh fading channel are considered. When the latency of both schemes is equal, the mismatched-decoding is generally better. However, when the complexity is equal (or where the codeword length of the underlying binary code is equal), the PBICM exponent is better in many cases.

The paper is organized as follows.

In Section II we review the classical BICM model and its properties, under the assumption of an infinite-length, ideal interleaver. In Section III the parallel BICM scheme is presented and the equivalence to a memoryless binary channel is established. In Section IV parallel BICM is studied from an information-theoretical viewpoint. Numerical examples and summary follow in Sections V and VI respectively.

## II. The BICM Communication Model

Notation: letters in bold ($\mathbf{x}, \mathbf{y}$...) denote row vectors, capital letters ($X, Y$...) denote random variables, and tilde denotes interleaved signals ($\tilde{\mathbf{b}}, \tilde{\mathbf{z}}$). $P_X(x)$ denotes the probability that the random variable (RV) $X$ will get the value $x$, and similarly $P_{Y|X}(y|x)$ denotes the probability $Y$ will get the value $y$ given that the RV $X$ is equal to $x$. $\mathbb{E}[\cdot]$ denotes statistical expectation. $\log$ means $\log_2$.

### A. Channel model

Let $W$ denote a memoryless channel with input and output alphabets $\mathcal{X}$ and $\mathcal{Y}$ respectively. The transition probabilities are defined by $W(y|x)$ for $y \in \mathcal{Y}$ and $x \in \mathcal{X}$. We assume that $\|\mathcal{X}\| = 2^L$. We consider equiprobable signaling only over the channel $W$.

An $(n, R)$ code $\mathcal{C} \subseteq \mathcal{X}^n$ is a set of $M = 2^{nR}$ codewords $\mathbf{c} \in \mathcal{X}^n$. The encoder wishes to convey one of $M$ equiprobable messages. The error probability of interest shall be the codeword error probability. An $(n, R)$ code with codeword error probability $p_e$ will sometimes be called an $(n, R, p_e)$ code.

### B. Classical BICM encoding and decoding

In BICM, a binary code is used to encode information messages $[m_1, m_2, ...]$ into binary codewords $[\mathbf{b}_1, \mathbf{b}_2, ...]$. The binary codewords are then interleaved using a long interleaver $\pi(\cdot)$, which applies a permutation on the coded bits. The interleaved bit stream $\tilde{\mathbf{b}}$ is partitioned into groups of $L$ consecutive bits and inserted into a mapper $\mu : \{0,1\}^L \to \mathcal{X}$. The mapper output, denoted $\mathbf{x}$, is fed into the channel. The encoding process is described in Figure 1.

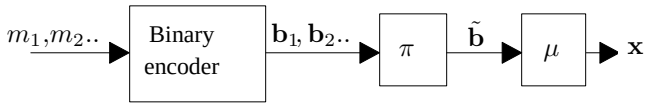


Fig. 1. BICM encoding process

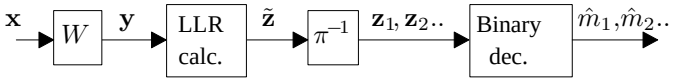


Fig. 2. BICM decoding process

The decoding process of BICM proceeds as follows. The channel output $\mathbf{y}$ is fed into a bit metric calculator, which calculates the log-likelihood ratio (LLR) of each input bit $b$ given the corresponding output sample $y$ ($L$ LLR values for each output sample). These LLR values (or bit metrics) denoted $\tilde{\mathbf{z}}$ are de-interleaved and partitioned into bit metrics $[\mathbf{z}_1, \mathbf{z}_2, ...]$ that correspond to the binary input codewords. Finally, the binary decoder decodes the messages $[\hat{m}_1, \hat{m}_2, ...]$ from $[\mathbf{z}_1, \mathbf{z}_2, ...]$. The decoding process is described in Figure 2.

The LLR of the $j^{th}$ bit in a symbol given the output value $y$ is calculated as follows:

$$LLR_j(y) \triangleq \log \frac{P_{Y|B_j}(y|0)}{P_{Y|B_j}(y|1)}, \tag{1}$$

where $P_{Y|B_j}(y|b)$ is the conditional probability of the channel output getting the value $y$ given that the $j^{th}$ bit at the mapper input was $b$, and the other $(L-1)$ bits are equiprobable independent binary random variables (RVs).

### *C. Classical BICM analysis: ideal interleaving*

In classical BICM (e.g. [2]) the LLR calculation is motivated by the assumption of a very long (*ideal*) interleaver $\pi$, so the coded bits go through essentially *independent* channels. These binary channels are defined as follows:

*Definition 1:* Let $W_i$ be a binary channel with transition probability

$$W_i(y|b) \triangleq \mathbb{E}\left[W(y|X = \mu(B_1, ..., B_L))|B_i = b\right] \tag{2}$$

$$= \frac{1}{2^{L-1}} \sum_{\substack{b_j;\ i\neq j \\ b_i = b}} W(y|\mu(b_1, ..., b_L)). \tag{3}$$

The channel $W_i(y|b_i)$ can be thought of as the original channel $W$ where the input is $x = \mu(b_1...b_L)$, where the bits $\{b_j\}_{j\neq i}$ are equiprobable independent RVs (see Fig. 3).

In [2], Caire et al. have proposed the following channel model for BICM called the *independent parallel channel model*. In this model the channel has a binary input $b$. A channel state $s$ is selected at random from $\mathcal{S} \triangleq \{1, ..., L\}$ with equal probability (and independently of $b$). Given a state $s$, the input bit $b$ is fed into the channel $W_s$. The channel outputs are the state $s$ and the output $y$ of the channel $W_s$. The channel, denoted by $\widetilde{W}$, is depicted in Figure 4.

The transition probability function of $\widetilde{W}$ is given by

$$\begin{aligned} \widetilde{W}(y, s|b) &= P_{Y,S|B}(y, s|b) \\ &= P_{Y|S,B}(y|s, b)P_S(s) \\ &= \frac{1}{L} W_s(y, b). \end{aligned} \tag{4}$$

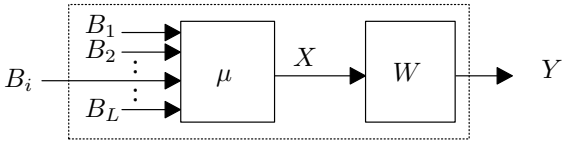


Fig. 3. The binary channel $W_i$. The bits $\{B_j\}_{j\neq i}$ are equiprobable independent RVs.

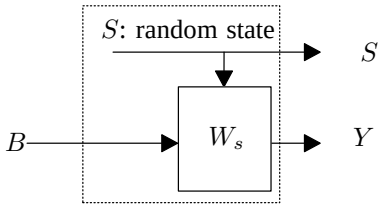


Fig. 4. The binary channel $\widetilde{W}$. The random state $S$ is known at the receiver.

Note that both outputs can be combined into a single output, the LLR, which is a sufficient statistic for optimal decoding over any binary-input channel. The LLR calculation for the channel $\widetilde{W}$ is given by

$$LLR_{\widetilde{W}}(y,s) = LLR_s(y), \tag{5}$$

where $LLR_s$ is given in (1).

Therefore the independent parallel channel model transforms the original nonbinary channel $W$ to a simple, memoryless channel. Using an infinite-length interleaver and a binary code that was designed for the simple binary channel $\widetilde{W}$, reliable communication for the original channel $W$ can be attained.

Let $\mathbf{C}(\cdot)$ denote the Shannon capacity of a channel (with equiprobable input).

*Lemma 1 (following [2]):* Let $\mathbf{C}^{\text{BICM}}(W)$ denote the capacity of the channel $W$ with BICM, a given mapping $\mu(\cdot)$ and an infinite-length interleaver (according to the independent parallel channel model). Denote by $\mathbf{C}(W_s)$ the capacity of the channel $W_s$. Then

$$\mathbf{C}^{\text{BICM}}(W) = \sum_{s=1}^{L} \mathbf{C}(W_s). \tag{6}$$

*Proof:* Since the independent parallel channel model assumes $L$ independent uses of the channel $\widetilde{W}$, we get that $\mathbf{C}^{\text{BICM}}(W) = L \cdot \mathbf{C}(\widetilde{W})$. The capacity of $\widetilde{W}$ is given by

$$\begin{aligned} \mathbf{C}\left(\widetilde{W}\right) &= I(B;Y,S) = I(B;Y|S) \\ &= \mathbb{E}_S I(B;Y|S=s) = \mathbb{E}_S \mathbf{C}(W_s) \\ &= \frac{1}{L}\sum_{s=1}^{L} \mathbf{C}(W_s). \end{aligned} \tag{7}$$

■

It is known that $\mathbf{C}^{\text{BICM}}(W)$ is generally smaller than the full channel capacity $\mathbf{C}(W)$, as opposed to other schemes, most notably multilevel coding and multistage decoding (MLC-MSD) [7], in which $\mathbf{C}(W)$ can be achieved. However, for Gray mapping the gap is small and can sometimes be tolerated. For example, for 8-PSK signaling over the AWGN channel with SNR = 5dB, $\mathbf{C}(W) = 1.86bit$ where $\mathbf{C}^{\text{BICM}}(W) = 1.84bit$.

## III. The Parallel BICM Scheme

In this section we propose an explicit BICM-type communication scheme which we call *parallel BICM* (PBICM), which allows the usage of binary codes on nonbinary channels at finite blocklength. The main features of the scheme include the following:

- Binary codewords are used *in parallel* to construct a codeword that enters the channel.
- A new finite-length interleaver.
- A random binary signal (binary dither) that is added to the binary codewords.

With the proposed scheme, we rigorously show how the original channel $W$ relates to the channel $\widetilde{W}$, thus allowing exact analysis and design of codes at finite block lengths.

### A. Interleaver Design

We wish to design a finite length interleaver, where:

- The length of the interleaver is minimal,
- The interleaver should be as simple as possible,
- The binary codewords will go through a binary memoryless channel.

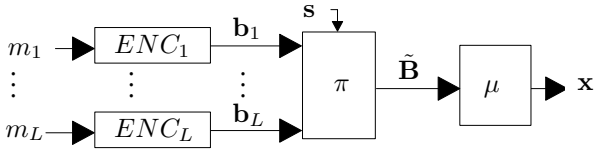


Fig. 5. Interleaving scheme viewed as parallel encoders

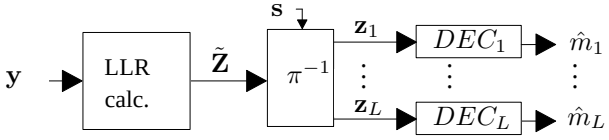


Fig. 6. De-interleaving scheme viewed as parallel decoders

In order for the binary codewords to experience a memoryless channel, each binary codeword must be spread over $n$ channel uses of $W$, so the interleaver output length cannot be less than $n$ channel uses. The newly proposed interleaver has of output length of exactly $n$, which satisfies the above requirements.

Let $ENC$ and $DEC$ be an encoder-decoder pair for a binary code. Let $\mathbf{b}_1, ..., \mathbf{b}_L$ be $L$ consecutive codewords from the output of $ENC$, bunched together to a matrix $\mathbf{B}$:

$$\mathbf{B} = \begin{pmatrix} \mathbf{b}_1 \\ \vdots \\ \mathbf{b}_L \end{pmatrix} = \begin{pmatrix} b_{11} & \dots & b_{1n} \\ \vdots & b_{lk} & \vdots \\ b_{L1} & \dots & b_{Ln} \end{pmatrix}. \tag{8}$$

Let $\mathbf{s}$ be a vector of i.i.d. random states drawn from $\mathcal{S}^n = \{1..L\}^n$. $\mathbf{s}$ shall be the interleaving signal. Each column in $\mathbf{B}$ shall be shifted cyclically by the corresponding element $s_k$, so the interleaved signal $\tilde{\mathbf{B}}$ is defined as

$$\tilde{\mathbf{B}} = \begin{pmatrix} b_{(1+s_1)_L 1} & \dots & b_{(1+s_n)_L n} \\ \vdots & b_{(l+s_k)_L k} & \vdots \\ b_{(L+s_1)_L 1} & \dots & b_{(1+s_n)_L n} \end{pmatrix},$$

where $(\xi)_L \triangleq (\xi \text{ modulo } L) + 1$.

Each column vector of interleaved signal $\tilde{\mathbf{B}}$ is mapped to a single channel symbol:

$$x_k = \mu(b_{(1+s_k)_L k}, \dots, b_{(L+s_k)_L k}), \tag{9}$$

and we call $\mathbf{x} = [x_1, ..., x_n]$ the *channel codeword*.

At the decoder an LLR value is calculated for every bit $b$ in $\tilde{\mathbf{B}}$ from $\mathbf{y}$. The LLR values are denoted by $\tilde{\mathbf{Z}}$. We assume that $\mathbf{s}$ is known at the decoder (utilizing common randomness), therefore the de-interleaving operation is simply sorting back the columns of $\tilde{\mathbf{Z}}$ according to $\mathbf{s}$ by reversing the modulo operation. The de-interleaver output is a vector of LLR values $\mathbf{z}$ for each transmitted codeword $\mathbf{b}$, according to (1). Each codeword is decoded independently by $DEC$.

### *B. Binary dither*

Since the decoder decodes each binary codeword independently, the communication scheme employing the above interleaver can be viewed as as set of parallel encoder-decoder pairs, which we denote by $ENC_1, ..., ENC_L$ and $DEC_1, ..., DEC_L$ (see Figures 5 and 6). We do not assume any independence between the effective channels between each encoder-decoder pair.

Consider the first encoder-decoder pair, $ENC_1$ and $DEC_1$. Since the input of $DEC_1$ depends on the codewords transmitted by $ENC_2$,...,$ENC_L$, the channel between $ENC_1$ and $DEC_1$ is not strictly memoryless. If, somehow, the decoders $DEC_2$,...,$DEC_L$ were forced to send i.i.d. equiprobable binary codewords, then the channel between $ENC_1$ and $DEC_1$ would be exactly the channel $\widetilde{W}$ (which is a binary memoryless channel) with the accurate LLR calculation (1).

In order to achieve the goal of $L$ binary memoryless channels between each encoder-decoder pair simultaneously, we add a binary dither - an i.i.d. equiprobable binary signal - to each encoder-decoder pair as follows.

Let the dither signals $\mathbf{d}_l = [d_{l1}, ..., d_{ln}]$, $l \in \{1, ..., L\}$ be $L$ random vectors, each of length $n$, that are drawn independently from a memoryless equiprobable binary source. The output of each encoder $ENC_l$, $\mathbf{b}_l$, goes through a component-wise XOR

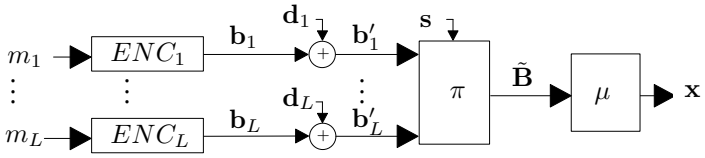


Fig. 7. PBICM encoding scheme. '$+$' denoted modulo-2 addition (XOR).

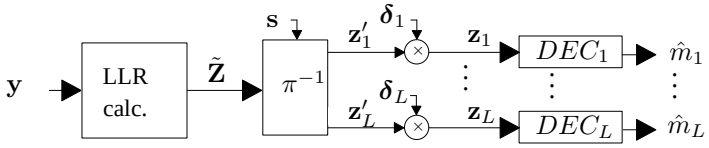


Fig. 8. PBICM decoding scheme. $\boldsymbol{\delta}_l \triangleq 1 - 2 \cdot \mathbf{d}_l$, '$\times$' denotes element-wise multiplication.

operation with the dither vector $\mathbf{d}_l$. The output of the XOR operation, denoted $\mathbf{b}'_l$, is fed into the interleaver $\pi$. The full PBICM encoding scheme is shown in Fig. 7.

We let each decoder $DEC_l$ know the value of the dither used by its corresponding encoder $ENC_l$, $\mathbf{d}_l$ (in practice the dither signals are generated using a pseudo-random generator which allows the common randomness). In order to compensate for the dither at the decoder, the LLR values are modified by flipping their sign for each dither value of 1 (and maintaining the sign where the dither is 0). Formally, denote the LLR values at the de-interleaver output by $\mathbf{z}'_l = [z'_{l1} \ ... \ z'_{ln}]$. The LLR values at the decoders input shall be denoted by $\mathbf{z}_l = [z_{l1} \ ... \ z_{ln}]$ and calculated as follows:

$$z_{lj} = z'_{lj}(1 - 2d_{lj}), \quad j = 1, ..., n. \tag{10}$$

The PBICM decoding scheme is shown in Fig. 8.

### C. Model equivalence

Before we analyze the channel between each encoder-decoder pair in PBICM, let us define a binary memoryless channel that is related to $\widetilde{W}$, that will prove useful in the analysis of PBICM.

*Definition 2:* Let $\overline{W}$ be a memoryless binary channel with input $B$ and output $\langle Y, S, D \rangle$: $S$ is drawn at random from $\{1, ..., L\}$, $D$ is drawn at random from $\{0, 1\}$ ($S$ and $D$ are independent, and both do not depend on the input $B$). $Y$ is the output of the channel $W_S$ with input $B \oplus D$ ($\oplus$ is the XOR operation). Note that the channel $\overline{W}$ is the channel $\widetilde{W}$ where the input is XORed with a binary RV $D$ (see Fig. 9).

Note that the LLR calculation for the channel $\overline{W}$ is given by

$$LLR_{\overline{W}}(y, s, d) = (-1)^d LLR_{\widetilde{W}}(y, s) = (-1)^d LLR_s(y), \tag{11}$$

where $LLR_{\widetilde{W}}$ and $LLR_s$ are given in (5) and (1), respectively.

*Theorem 1:* In parallel BICM, the channel between every encoder-decoder pair is exactly the binary memoryless channel $\overline{W}$, with its exact LLR output.

*Proof:* Consider the pair $ENC_1$ and $DEC_1$. Let $\mathbf{b}_1$ be the codeword sent from $ENC_1$. After adding the dither $\mathbf{d}_1$, the dithered codeword $\mathbf{b}'_1$ enters the interleaver. The other codewords $\mathbf{b}_2, ..., \mathbf{b}_L$ are dithered using $\mathbf{d}_2, ..., \mathbf{d}_L$. Since the dither of these codewords is unknown at $DEC_1$, the dithered codewords $\mathbf{b}'_2, ..., \mathbf{b}'_L$ are truly random i.i.d. signals. The interleaving signal $\mathbf{s}$ interleaves the dithered codewords according to (8). The interleaved signal enters the mapper $\mu$ and the channel $W$,

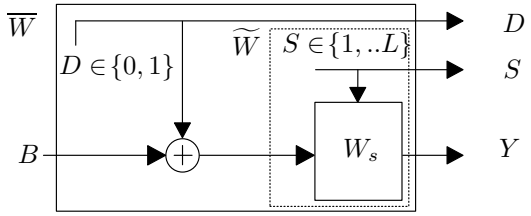


Fig. 9. The binary channel $\overline{W}$. The random state $S$ and the dither $D$ are known at the receiver.

resulting in an output $\mathbf{y}$. Since the dithered codewords $\mathbf{b}'_2, ..., \mathbf{b}'_L$ are i.i.d., the equivalent channel from $\mathbf{b}'_1$ to $\langle \mathbf{y}, \mathbf{s} \rangle$ is exactly the channel $\widetilde{W}$. The LLR calculation at the PBICM receiver along with the interleaver produce $\mathbf{z}'_1$ , which is exactly the LLR calculation that fits the channel $\widetilde{W}$ (cf. (5)).

Recalling that the channel $\overline{W}$ is nothing but the channel $\widetilde{W}$ with its input XORed with a binary RV, and that the LLR of the channel $\widetilde{W}$ can be easily modified by the dither according to Eq. (11) to produce the LLR of the channel $\overline{W}$, we conclude that the channel between $\mathbf{b}_1$ to $\mathbf{z}_1$ is exactly the channel $\overline{W}$ with $LLR$ calculation.

Since by symmetry the above holds for any encoder-decoder pair $ENC_l$-$DEC_l$, the proof is concluded. ■

An important note should be made: Parallel BICM allows the decomposition of the nonbinary channel $W$ to $L$ binary channels of the type $\overline{W}$. These $L$ channels are *not* independent. For example, if $W$ is an additive noise channel, and at some point the noise instance is very strong, this will affect all the decoders and they will fail in decoding together. However, since in the PBICM scheme the channels are used independently, the operation of each decoder depends only on the marginal distribution of the relevant channel outputs. The outputs of these decoders will inevitably be statistically dependent, and we take this into consideration when analyzing the performance of coding using PBICM in the following.

### *D. Error Probability Analysis*

We wish to analyze the performance of PBICM, and specifically, we are interested in the overall codeword error probability. Let $\mathcal{C}$ be a binary $(n, R)$ code, used in the PBICM scheme. To assure a fair comparison, we regard each $L$ consecutive information messages $(m_1, ..., m_L)$ as a single message $m$, and regard the scheme as a code of length $n$ on the channel input alphabet $\mathcal{X}$. We define the following error events: Let $\mathcal{E}_l$ be the event of a codeword error in $DEC_l$, and let $\mathcal{E}$ be the event of an error in *any* of the messages $\{m_1, ..., m_L\}$, i.e. $\mathcal{E} = \bigcup_l \mathcal{E}_l$. Denote the corresponding error probabilities by $p_{e_l}$ and $p_e$ respectively.

*Corollary 1:* Let $p_e(\overline{W})$ be the codeword error probability of the code $\mathcal{C}$ over the channel $\overline{W}$. Then the overall error probability $p_e$ of the code $\mathcal{C}$ used with PBICM can be bounded by

$$p_e(\overline{W}) \leq p_e \leq L \cdot p_e(\overline{W}). \tag{12}$$

*Proof:* Since the error events $\mathcal{E}_l$ in codewords that are mapped to the same channel codeword together are dependent, we can only bound the overall error probability $p_e$ using the union bound. $p_e$ can be also lower bounded by the minimum of the error probabilities in any of the channels:

$$\min\{p_{e_1}, ..., p_{e_L}\} \leq p_e \leq \sum_l p_{e_l}. \tag{13}$$

Since by Theorem 1 the channel between each of the encoder-decoder pairs is $\overline{W}$, we get that the error probabilities must be all equal to the error probability of the code $\mathcal{C}$ over the channel $\overline{W}$. Setting $p_{e_1} = p_{e_2} = ... = p_{e_L} = p_e(\overline{W})$ in (13) completes the proof. ■

In many cases the bit error rate (BER) is of interest. Suppose that each of the messages $(m_1, ..., m_L)$ represents $k$ information bits and the entire message $m$ represents $L \cdot k$ information bits. Let $\mathcal{E}^b_{lk'}$ denote the error in the $k'$-th bit of the information message $m_l$. The average BER for the encoder-decoder pair $ENC_l$-$DEC_l$ is defined by

$$p^b_{e_l} \triangleq \frac{1}{k} \sum_{k'=1}^{k} Pr\{\mathcal{E}^b_{lk'}\}. \tag{14}$$

Similarly, define the overall average BER as

$$p^b_e \triangleq \frac{1}{L \cdot k} \sum_{l=1}^{L} \sum_{k'=1}^{k} Pr\{\mathcal{E}^b_{lk'}\} = \frac{1}{L} \sum_{l=1}^{L} p^b_{e_l}. \tag{15}$$

*Corollary 2:* Let $p^b_e(\overline{W})$ be the average BER of a binary code $\mathcal{C}$ over the channel $\overline{W}$. Then the average BER $p^b_e$ of the code $\mathcal{C}$ used with PBICM is equal to $p^b_e(\overline{W})$.

*Proof:* Follows directly from Theorem 1 and from the definition of the average BER in (15). ■

## IV. Parallel BICM: Information Theoretical Analysis

In the previous section we defined the PBICM scheme and analyzed its basic error probability properties. The equivalence of the channel between each encoder-decoder pair that was established in Theorem 1 enables a full information-theoretical analysis of the scheme. We show that the highest achievable rate by PBICM (the PBICM capacity) is equal to the BICM capacity as in Equation (7), which should not be a surprise. At the finite-length regime, we derive error exponent and channel dispersion results as information-theoretical measures for optimal PBICM performance at finite-length.

### A. Capacity

Let the PBICM capacity of $W$, $\mathbf{C}^{\text{PBICM}}(W)$, be the highest achievable rate for reliable communication over the channel $W$ with PBICM and a given mapping $\mu$. (As usual, reliable communication means a vanishing codeword error probability as the codelength $n$ goes to infinity.)

*Theorem 2:* The PBICM capacity is given by

$$\mathbf{C}^{\text{PBICM}}(W) = L \cdot \mathbf{C}(\overline{W}) = \sum_{s=1}^{L} \mathbf{C}(W_s) = \mathbf{C}^{BICM}(W). \tag{16}$$

*Proof:*

*Achievability:* Let $\mathcal{C}^{(n)}$ be a series of (binary) capacity-achieving codes for the channel $\overline{W}$, and let $p_e^{(n)}(\overline{W})$ be the corresponding (vanishing) codeword error probabilities. By Corollary 1, the overall error probability of PBICM with a binary code is upper bounded by $L$ times the error probability of the same code over the channel $\overline{W}$, therefore when the codes $\mathcal{C}^{(n)}$ are used with PBICM, the overall error probability is bounded by $L \cdot p_e^{(n)}(\overline{W})$ and also vanish with $n$. Since there are $L$ instances of the channel $\overline{W}$, we get that the rate of $L \cdot \mathbf{C}(\overline{W})$ is achievable by PBICM.

*Converse:* Let $\mathcal{C}^{(n)}$ be a series of binary codes that are used with PBICM and achieve a vanishing overall error probability $p_e^{(n)}$, and suppose that the overall PBICM rate is given by $L \cdot R$ (a rate of $R$ at each encoder-decoder pair). By Corollary 1, the codeword error probability of a code over $\overline{W}$ is upper bounded by the overall error probability of the same code used in PBICM. Therefore, if $p_e^{(n)}$ vanishes as $n \to \infty$, then the error probability over $\overline{W}$ must also vanish, and therefore the communication rate between each encoder-decoder pair must be upper bounded by $\mathbf{C}(\overline{W})$, and the overall rate cannot surpass $L \cdot \mathbf{C}(\overline{W})$.

All that remains is to calculate the capacity of $\overline{W}$:

$$\mathbf{C}(\overline{W}) = I(B;Y,S,D) = I(B;Y,S|D) = \frac{1}{2}\left(I(B;Y,S|D=0) + I(B;Y,S|D=1)\right). \tag{17}$$

When $D=0$, we get the channel $\widetilde{W}$ exactly, and when $D=1$ we get the channel $\widetilde{W}$ with its input symbols always switched. In either way, the expression $I(B;Y,S|D=d)$ is equal to the capacity of $\widetilde{W}$. Using Lemma 1, we get that

$$\mathbf{C}(\overline{W}) = \mathbf{C}(\widetilde{W}) = \frac{1}{L}\sum_{s=1}^{L} \mathbf{C}(W_s). \tag{18}$$

■

A note regarding the capacity proof: one might me tempted to try and prove the capacity theorem for PBICM without dither, since with random coding, the code $\mathcal{C}$ is merely an i.i.d. binary random vector. This approach fails because of the following. In the decoding of each codeword, the correctness of the model $\widetilde{W}$ relies on the fact that the *other* codewords are i.i.d. signals. Since PBICM requires a single code for all the $L$ levels, such a condition can never be met. It it possible to prove the achievability without dither when using a different random code at each level, but such an approach will not guarantee the existence of a single code, as required by PBICM.

### B. Error Exponent

The error exponent of a channel $W$ is defined by

$$\mathbf{E}(R) \triangleq \lim_{n\to\infty} -\frac{1}{n}\log\left(p_e(n)\right), \tag{19}$$

where $p_e(n)$ is the average codeword error probability for the best code of length $n$. A lower bound on the error exponent for memoryless channels is the *random coding* error exponent [3], which is given by

$$\mathbf{E}_r(R) = \max_{\rho\in[0,1]} \max_{P_X(\cdot)} \{\mathbf{E}_0(\rho, P_X) - \rho R\}, \tag{20}$$

where

$$\mathbf{E}_0(\rho, P_X) \triangleq -\log\left[\sum_{y\in\mathcal{Y}}\left(\sum_{x\in\mathcal{X}} P_X(x)W(y|x)^{1/(1+\rho)}\right)^{1+\rho}\right]. \tag{21}$$

Since we consider equiprobable inputs only we omit the dependence of $\mathbf{E}_0(\rho)$ in $P_X$, and omit the maximization w.r.t. $P_X$ in (20).

Others known bounds on the error exponent include the *expurgation* error exponent lower bound, the *sphere packing* error exponent (an upper bound) and others [3]. The random coding and sphere packing exponents coincide for rates above the critical rate, and therefore the error exponent is known precisely at these rates.

*1) PBICM error exponent:*

Similarly to (19), we define the PBICM error exponent:

*Definition 3:* For a given channel $W$ and a mapping $\mu$, let $\mathbf{E}^{\text{PBICM}}(R)$ be defined as

$$\mathbf{E}^{\text{PBICM}}(R) \triangleq \lim_{n\to\infty} -\frac{1}{n}\log\left(p_e(n)\right), \tag{22}$$

where $p_e(n)$ is the average codeword error probability for the best PBICM scheme with block length of $n$.

Using Corollary 1, we can calculate the PBICM exponent using the error exponent of $\overline{W}$:

*Theorem 3:* The PBICM error exponent of a channel $W$ is given by

$$\mathbf{E}^{\text{PBICM}}(R) = \mathbf{E}(R/L), \tag{23}$$

where $\mathbf{E}(\cdot)$ is the error exponent function of the binary channel $\overline{W}$.

*Proof:* Let $\mathcal{C}^{(n)}$ be a series of the binary codes. Denote their codeword error probabilities over the channel $\overline{W}$ by $p_e^{(n)}(\overline{W})$. Let $p_e^{(n)}$ be the error probabilities of the corresponding PBICM schemes with $\mathcal{C}^{(n)}$ used as underlying codes. It follows from (12) that

$$-\frac{1}{n}\log(L\cdot p_e^{(n)}(\overline{W})) \le -\frac{1}{n}\log p_e^{(n)} \le -\frac{1}{n}\log p_e^{(n)}(\overline{W}). \tag{24}$$

By taking $n\to\infty$ the factor of $L$ vanishes and we get that for any series of codes,

$$\lim_{n\to\infty} -\frac{1}{n}\log p_e^{(n)} = \lim_{n\to\infty} -\frac{1}{n}\log p_e^{(n)}(\overline{W}). \tag{25}$$

The above equation holds for the series of best codes for the channel $\overline{W}$, as well as for the series of the best codes for PBICM. Therefore the equality holds for the sequence of best codes on either side. Since the rate for PBICM is $L$ times the rate for coding on $\overline{W}$, the proof is concluded. ■

*2) The error exponent of $\overline{W}$:*

The channel $\overline{W}$ has a special structure, and is related to the binary sub-channels $W_i$. We now calculate two basic bounds for the error exponent of $\overline{W}$ in terms of the sub-channels $W_i$. By Theorem 3, the PBICM error exponent of the channel $W$ can be bounded accordingly.

*Theorem 4:* Let $\mathbf{E}(R)$ be the error exponent of the channel $\overline{W}$. It can be bounded as follows:

*Random coding:*

$$\mathbf{E}(R) \ge \mathbf{E}_r(R) = \max_{\rho\in[0,1]}\{\mathbf{E}_0(\rho) - \rho R\}, \tag{26}$$

where

$$\mathbf{E}_0(\rho) = -\log\mathbb{E}\left[2^{-\mathbf{E}_0^{(S)}(\rho)}\right], \tag{27}$$

$\mathbf{E}_0^{(s)}(\rho)$ is the $\mathbf{E}_0$ function for the channel $W_s$, and the expectation is w.r.t. the state $S$ which is drawn uniformly from $\{1..L\}$.

*Sphere packing:*

$$\mathbf{E}(R) \le \mathbf{E}_{sp}(R) = \max_{\rho>0}\{\mathbf{E}_0(\rho) - \rho R\}, \tag{28}$$

where $\mathbf{E}_0(\rho)$ is given in (27).

*Proof:*

The bounds in the theorem are the original random coding and sphere packing exponents [3]. The proof, therefore, boils down to the simplification of the $\mathbf{E}_0$ function to the form of (27).

Consider the channel $\overline{W}$ (Definition 2) with binary input $B$ and outputs $\langle Y, S, D\rangle$. Since $\overline{W}$ is equivalent to the channel $\widetilde{W}$ with input $B\oplus D$, where $D$ is an equiprobable binary RV (and known at the receiver), we get that

$$\overline{W}(y,s,d|b) = \frac{1}{2}\widetilde{W}(y,s|b\oplus d). \tag{29}$$

The channel $\widetilde{W}$, in turn, is nothing more than the channel $W_s$ with the additional output $S$. This yields

$$\frac{1}{2}\widetilde{W}(y,s|b\oplus d) = \frac{1}{2L}W_s(y|b\oplus d). \tag{30}$$

Combining the above, the function $\mathbf{E}_0$ of $\overline{W}$ is therefore given by

$$\begin{aligned}
\mathbf{E}_0(\rho) &= -\log \sum_{\substack{y\in\mathcal{Y}\\ s\in\{1..L\}\\ d\in\{0,1\}}} \left[\sum_{b\in\{0,1\}} P_B(b)\overline{W}(y,s,d|b)^{\frac{1}{1+\rho}}\right]^{1+\rho} \\
&= -\log \sum_{\substack{y\in\mathcal{Y}\\ s\in\{1..L\}\\ d\in\{0,1\}}} \left[\sum_{b\in\{0,1\}} \frac{1}{2}\left(\frac{1}{2L}W_s(y|b\oplus d)\right)^{\frac{1}{1+\rho}}\right]^{1+\rho} \\
&\overset{(a)}{=} -\log \sum_{\substack{y\in\mathcal{Y}\\ s\in\{1..L\}}} \left[\sum_{b'\in\{0,1\}} \frac{1}{2}\left(\frac{1}{L}W_s(y|b')\right)^{\frac{1}{1+\rho}}\right]^{1+\rho} \qquad (31) \\
&= -\log \sum_{s\in\{1..L\}} \frac{1}{L}\sum_{y\in\mathcal{Y}}\left[\sum_{b'\in\{0,1\}} \frac{1}{2}W_s(y|b')^{\frac{1}{1+\rho}}\right]^{1+\rho} \\
&\overset{(b)}{=} -\log \sum_{s\in\{1..L\}} \frac{1}{L}2^{-\mathbf{E}_0^{(s)}(\rho)} \\
&= -\log \mathbb{E}\left[2^{-\mathbf{E}_0^{(s)}(\rho)}\right] \qquad (32)
\end{aligned}$$

$(a)$ follows by setting $b' \triangleq b\oplus d$, and by noting that the summation result is independent of the value of $d$. $(b)$ follows from the definition of $\mathbf{E}_0^{(s)}(\rho)$ (the $\mathbf{E}_0$ function for the channel $W_s$).

■

Several notes can be made:

- It is well known that the random coding and sphere packing exponents coincide at rates above the critical rate. Therefore the exact error exponent of $\overline{W}$ is known at rates above the critical rate of $\overline{W}$, $R_{cr}^{\overline{W}}$. It follows that the exact PBICM error exponent is known at rates above $R_{cr}^{\text{PBICM}} \triangleq L\cdot R_{cr}^{\overline{W}}$, which we define to be the PBICM critical rate.
- In theorem 4 we have shown that the random coding and the sphere packing bounds have a compact form because of the special structure of the channel $\overline{W}$. Clearly, following Theorem 3, every bound on $\mathbf{E}(R)$ of $\overline{W}$ serves as a bound on the PBICM error exponent. However, for other bounds (such as the expurgation error exponent [3]), no compact form could be found. Such bounds, of course, can still be applied to bound $\mathbf{E}^{\text{PBICM}}(R)$.
- The $\mathbf{E}_0$ function of the channel $\widetilde{W}$ is equal to the $\mathbf{E}_0$ function of the channel $\overline{W}$. This can easily be seen from the proof above: $\mathbf{E}_0$ for $\widetilde{W}$ is given in (31) by definition.
- In [7], the authors offered the model of $\widetilde{W}$ for calculating the error exponent of BICM. It is claimed that $\mathbf{E}_0$ of the channel $\widetilde{W}$ is given by [7, Eq. (37)]:

$$\mathbb{E}\left[\mathbf{E}_0^{(S)}(\rho)\right] = \frac{1}{L}\sum_{s=1}^{L}\mathbf{E}_0^{(s)}(\rho). \qquad (33)$$

  As we have just shown in Theorem 4, this is not the exact expression. In fact, it can be shown that $\mathbf{E}_0(\rho) \le \mathbb{E}\left[\mathbf{E}_0^{(S)}(\rho)\right]$. This follows directly from the convexity of the function $2^{-(\cdot)}$ and the Jensen inequality. Therefore the incorrect expression in [7, Eq. (37)] always overestimates the value of $\mathbf{E}_0(\rho)$, and therefore the resulting $\mathbf{E}_r(R)$ expression also overestimates the true random coding expression.

### C. Channel Dispersion

An alternative information theoretical measure for quantifying coding performance with finite block lengths is the *channel dispersion*. Suppose that a fixed codeword error probability $p_e$ and a codeword length $n$ are given. We can then seek the maximal achievable rate $R$ given $p_e$ and $n$.

It appears that for fixed $p_e$ and $n$, the gap to the channel capacity is approximately proportional to $Q^{-1}(p_e)/\sqrt{n}$ (where $Q(\cdot)$ is the complementary Gaussian cumulative distribution function). The proportion constant (squared) is called the channel dispersion. Formally, define the (operational) channel dispersion as follows [6]:

*Definition 4:* The dispersion $\mathbf{V}(W)$ of a channel $W$ with capacity $C$ is defined as

$$\mathbf{V}(W) = \lim_{p_e\to 0}\limsup_{n\to\infty}\; n\cdot\left(\frac{C-R(n,p_e)}{Q^{-1}(p_e)}\right)^2, \qquad (34)$$

where $R(n,p_e)$ is the highest achievable rate for codeword error probability $p_e$ and codeword length $n$.

In 1962 , Strassen [4] used the Gaussian approximation to derive the following result for DMCs[1]:

$$R(n,p_e) = C - \sqrt{V/n}Q^{-1}(p_e) + O\left(\frac{\log n}{n}\right), \tag{35}$$

where $C$ is the channel capacity, and the new quantity $V$ is the (information-theoretic) dispersion , which is given by

$$V \triangleq VAR(i(X;Y)), \tag{36}$$

where $i(x;y)$ is the information spectrum, given by

$$i(x;y) \triangleq \log \frac{P_{XY}(x,y)}{P_X(x)P_Y(y)}, \tag{37}$$

and the distribution of $X$ is the capacity-achieving distribution that minimizes $V$. Strassen's result proves that the dispersion of DMCs is equal to $VAR(i(X;Y))$. This result was recently tightened (and extended to the power-constrained AWGN channel) in [6]. It is also known that the channel dispersion and the error exponent are related as follows. For a channel with capacity $C$ and dispersion $V$, the error exponent can be approximated by $\mathbf{E}(R) \cong \frac{(C-R)^2}{2V \ln 2}$. See [6] for details on the early origins of this approximation by Shannon.

*1) PBICM dispersion:* In order to estimate the finite-block performance of PBICM schemes we extend the dispersion definition as follows:

*Definition 5:* The PBICM dispersion $\mathbf{V}^{\text{PBICM}}(W)$ of a channel $W$ with a given mapping $\mu$ and PBICM capacity $\mathbf{C}^{\text{PBICM}}(W)$ is defined as

$$\mathbf{V}^{\text{PBICM}}(W) = \lim_{p_e \to 0} \limsup_{n\to\infty} \; n \cdot \left(\frac{\mathbf{C}^{\text{PBICM}}(W) - R(n,p_e)}{Q^{-1}(p_e)}\right)^2, \tag{38}$$

where $R(n,p_e)$ is the highest achievable rate for any PBICM scheme with a given $n$ and $p_e$.

Relying on the relationship between the PBICM scheme and the binary channel $\overline{W}$, we can show the following:

*Theorem 5:* Let $n$ be a given block length and let $p_e$ be a given codeword error probability. The highest achievable rate attained using PBICM, $R^{\text{PBICM}}(n,p_e)$ is bounded from above and below by:

$$R^{\text{PBICM}}(n,p_e) \;\geq\; \mathbf{C}^{\text{PBICM}}(W) - \sqrt{\frac{L^2\mathbf{V}(\overline{W})}{n}}Q^{-1}\left(\frac{p_e}{L}\right) + O\left(\frac{1}{n}\right), \tag{39}$$

$$R^{\text{PBICM}}(n,p_e) \;\leq\; \mathbf{C}^{\text{PBICM}}(W) - \sqrt{\frac{L^2\mathbf{V}(\overline{W})}{n}}Q^{-1}(p_e) + O\left(\frac{\log n}{n}\right). \tag{40}$$

As a result, the PBICM dispersion is given by

$$\mathbf{V}^{\text{PBICM}}(W) = L^2\mathbf{V}(\overline{W}). \tag{41}$$

*Proof:*

*Direct:* From the achievability proof of (35) [6, Theorem 45], there must exist an $(n, R', p'_e = p_e/L)$ binary code for $\overline{W}$ that satisfies

$$R' \geq \mathbf{C}(\overline{W}) - \sqrt{\frac{\mathbf{V}(\overline{W})}{n}}Q^{-1}(p_e/L) + O\left(\frac{1}{n}\right). \tag{42}$$

By Theorem 1 and Corollary 1, it follows that the PBICM scheme based on this code is not greater than $Lp'_e = p_e$. The rate of the PBICM scheme satisfies

$$R \;=\; L\cdot R' \geq L\left[\mathbf{C}(\overline{W}) - \sqrt{\frac{\mathbf{V}(\overline{W})}{n}}Q^{-1}(p'_e) + O\left(\frac{1}{n}\right)\right] \tag{43}$$

$$\;=\; \mathbf{C}^{BICM}(W) - \sqrt{\frac{L^2\mathbf{V}(\overline{W})}{n}}Q^{-1}\left(\frac{p_e}{L}\right) + O\left(\frac{1}{n}\right). \tag{44}$$

[1]see Appendix B for the big-O notation.

*Converse:* Suppose we have a $(n, R, p_e)$ PBICM scheme. According to Corollary 1, the codeword error probability $p'_e$ of the underlying binary code is not greater than than $p_e$. By Equation (35), the rate $R'$ of the underlying binary code is bounded by

$$R' \leq \mathbf{C}(\overline{W}) - \sqrt{\frac{\mathbf{V}(\overline{W})}{n}} Q^{-1}(p'_e) + O\left(\frac{\log n}{n}\right). \tag{45}$$

Since $Q^{-1}(\cdot)$ is a decreasing function, the bound loosens by replacing $p'_e$ with the higher $p_e$. Therefore the overall rate $R$ is bounded by

$$\begin{aligned} R &= L \cdot R' \leq L\left[\mathbf{C}(\overline{W}) - \sqrt{\frac{\mathbf{V}(\overline{W})}{n}} Q^{-1}(p'_e) + O\left(\frac{\log n}{n}\right)\right] &(46)\\ &\leq \mathbf{C}^{\text{PBICM}}(W) - \sqrt{\frac{L^2\mathbf{V}(\overline{W})}{n}} Q^{-1}\left(p_e\right) + O\left(\frac{\log n}{n}\right). &(47)\end{aligned}$$

*PBICM dispersion:* Rewriting Equations (39) and (40), we get the following:

$$\sqrt{\frac{L^2\mathbf{V}(\overline{W})}{n}} Q^{-1}(p_e) + O\left(\frac{\log n}{n}\right) \leq \mathbf{C}^{\text{PBICM}}(W) - R \leq \sqrt{\frac{L^2\mathbf{V}(\overline{W})}{n}} Q^{-1}\left(\frac{p_e}{L}\right) + O\left(\frac{1}{n}\right) \tag{48}$$

$$\sqrt{L^2\mathbf{V}(\overline{W})} + O\left(\frac{\log n}{\sqrt{n}}\right) \leq \sqrt{n}\left(\frac{\mathbf{C}^{\text{PBICM}}(W) - R}{Q^{-1}(p_e)}\right) \leq \sqrt{L^2\mathbf{V}(\overline{W})} \cdot \frac{Q^{-1}\left(\frac{p_e}{L}\right)}{Q^{-1}(p_e)} + O\left(\frac{1}{\sqrt{n}}\right) \tag{49}$$

Taking the limit w.r.t. $n$ yields

$$\sqrt{L^2\mathbf{V}(\overline{W})} \leq \limsup_{n\to\infty} \sqrt{n}\left(\frac{\mathbf{C}^{\text{PBICM}}(W) - R}{Q^{-1}(p_e)}\right) \leq \sqrt{L^2\mathbf{V}(\overline{W})} \cdot \frac{Q^{-1}\left(\frac{p_e}{L}\right)}{Q^{-1}(p_e)}, \tag{50}$$

or

$$L^2\mathbf{V}(\overline{W}) \leq \limsup_{n\to\infty} n\left(\frac{\mathbf{C}^{\text{PBICM}}(W) - R}{Q^{-1}(p_e)}\right)^2 \leq L^2\mathbf{V}(\overline{W})\left(\frac{Q^{-1}\left(\frac{p_e}{L}\right)}{Q^{-1}(p_e)}\right)^2. \tag{51}$$

By noting that $\lim_{\varepsilon\to 0^+} \frac{Q^{-1}(\varepsilon)^2}{2\ln\frac{1}{\varepsilon}} = 1$ (see Appendix A), we get that

$$\lim_{p_e\to 0}\left(\frac{Q^{-1}\left(\frac{p_e}{L}\right)}{Q^{-1}(p_e)}\right)^2 = \lim_{p_e\to 0}\frac{\ln(L/p_e)}{\ln(1/p_e)} = \lim_{p_e\to 0}\frac{\ln(1/p_e) + \ln L}{\ln(1/p_e)} = 1, \tag{52}$$

which leads to the desired result:

$$\mathbf{V}^{\text{PBICM}}(W) = \lim_{p_e\to 0}\limsup_{n\to\infty}\ n \cdot \left(\frac{\mathbf{C}^{\text{PBICM}}(W) - R(n, p_e)}{Q^{-1}(p_e)}\right)^2 = L^2\mathbf{V}(\overline{W}). \tag{53}$$

■

Note that the PBICM dispersion result is not as tight as the bound for general coding schemes as in (35). The reason is the unavoidable use of the union bound when estimating the overall error probability of PBICM in Theorem 1. In the dispersion proof for DMCs, the value of the dispersion is obtained even without taking the limit w.r.t. $p_e$. However, the gap between $Q^{-1}(p_e)$ and $Q^{-1}(p_e/L)$ for values of interest is not very large.

*2) The dispersion of $\overline{W}$:*

As in the error exponent case, the PBICM dispersion of a channel is related to the dispersion of the binary channel $\overline{W}$. We now calculate it explicitly from the dispersions of the sub-channels $W_i$.

*Theorem 6:* The dispersion of the channel $\overline{W}$ is given by

$$\mathbf{V}(\overline{W}) = \mathbf{V}(\widetilde{W}) = \mathbb{E}[\mathbf{V}(W_S)] + VAR\left[\mathbf{C}(W_S)\right] = \left[\frac{1}{L}\sum_{s=1}^{L}\mathbf{V}(W_s)\right] + VAR(\mathbf{C}(W_S)) \tag{54}$$

where $VAR(\mathbf{C}(W_S))$ is the statistical variance of the capacity of $W_s$, i.e.

$$VAR(\mathbf{C}(W_S)) \triangleq \mathbb{E}[\mathbf{C}^2(W_S)] - \mathbb{E}^2[\mathbf{C}(W_S)]. \tag{55}$$

*Proof:* Consider the channel $\overline{W}$ (Definition 2) with binary input $B$ and outputs $\langle Y, S, D\rangle$, and recall that

$$P_{YSD|B}(y,s,d|b) = \overline{W}(y,s,d|b) = \frac{1}{2}\widetilde{W}(y,s|b\oplus d) = \frac{1}{2L}W_s(y|b\oplus d). \tag{56}$$

We first calculate the dispersion of $\widetilde{W}$. Since $S$ and the channel input $B$ are independent, the information spectrum is given by

$$\begin{aligned} i(b;y,s) &\triangleq \log\frac{P_{YSB}(y,s,b)}{P_{YS}(y,s)P_B(b)} = \log\frac{P_{Y|SB}(y|s,b)P_S(s)P_B(b)}{P_{YS}(y,s)P_B(b)} &(57)\\ &= \log\frac{P_{Y|SB}(y|s,b)}{P_{Y|S}(y|s)} \triangleq i(b;y|s). &(58)\end{aligned}$$

Using this notation, the dispersion of the channel $W_s$ is given by

$$\begin{aligned}\mathbf{V}(W_s) &= VAR(i(B;Y|s)|S=s)\\ &= \mathbb{E}\left[i^2(B;Y|s)|S=s\right] - \mathbf{C}(W_s)^2.\end{aligned}$$

Next, the dispersion of the channel $\widetilde{W}$ is given as follows:

$$\begin{aligned}\mathbf{V}(\widetilde{W}) &= VAR(i(B;Y,S)) = VAR(i(B;Y|S))\\ &\overset{(a)}{=} \mathbb{E}\left[VAR[i(B;Y|s)|S=s]\right] + VAR\left[\mathbb{E}[i(B;Y|S)|S=s]\right]\\ &= \mathbb{E}[\mathbf{V}(W_S)] + VAR\left[\mathbf{C}(W_S)\right] = \left[\frac{1}{L}\sum_{s=1}^{L}\mathbf{V}(W_s)\right] + VAR(\mathbf{C}(W_S)). &(59)\end{aligned}$$

$(a)$ follows from the law of total variance.

Finally, the dispersion of the channel $\overline{W}$ is calculated as follows:

Let us combine the outputs of the channel $\widetilde{W}$ to a single output $Z = \langle Y, S\rangle$. We therefore end up with a channel with input $B$ and outputs $Z$ and $D$ (see Fig. 9). Similarly to (57), we get that the information spectrum is given by

$$i(b;z,d) \triangleq \log\frac{P_{ZDB}(z,d,b)}{P_{ZD}(z,d)P_B(b)} = i(b;z|d). \tag{60}$$

Following (59), we get that

$$\mathbf{V}(\overline{W}) = \mathbb{E}[\mathbf{V}(\widetilde{W}_D)] + VAR\left[\mathbf{C}(\widetilde{W}_D)\right] = \frac{1}{2}\sum_{d=\{0,1\}}\mathbf{V}(\widetilde{W}_d) + VAR(\mathbf{C}(\widetilde{W}_D)), \tag{61}$$

where $\widetilde{W}_d$ is the channel $\widetilde{W}$ with its input XORed with the value $d$.

Since only equiprobable inputs are considered, it follows that $\mathbf{C}(\widetilde{W}_0) = \mathbf{C}(\widetilde{W}_1) = \mathbf{C}(\widetilde{W})$, and that $\mathbf{V}(\widetilde{W}_0) = \mathbf{V}(\widetilde{W}_1) = \mathbf{V}(\widetilde{W})$. It therefore follows that $VAR(\mathbf{C}(\widetilde{W}_D)) = 0$, and consequently, $\mathbf{V}(\overline{W}) = \mathbf{V}(\widetilde{W})$, as required. ■

Note that since large dispersion means higher backoff from the capacity (see (35)), the term $VAR(\mathbf{C}(W_S))$ can be thought of as a *penalty factor* for the dispersion, over the expected dispersion over the channels $W_s$, $\mathbb{E}[\mathbf{V}(W_S)]$. This factor grows as the capacities of the sub-channels $W_i$ are more spread.

## V. Numerical Results

In this section we evaluate numerically the information-theoretical quantities for PBICM. In particular, we calculate the PBICM random coding error exponent (see Theorems 3 and 4) in order to compare with the mismatched decoding approach [8]. We consider the AWGN channel and the Rayleigh fading channel (with perfect channel state information at the receiver) over a wide range of SNR values and constellations. Gray mapping was used throughout all the examples.

### *A. Normalization: latency vs. complexity*

One way to compare the PBICM error exponent with the mismatched decoding exponent is to compare the error probability when the block length $n$ is fixed, which gives a simple comparison between the exponent values. Such an approach makes sense, since both schemes have the same latency of $n$ channel uses. As will be seen in the coming examples, for fixed $n$ the PBICM error exponent is inferior to that of the mismatched decoding. This can also be seen by observing that the PBICM random coding exponent has a slope of $-1/L$ (in its straight-line region), where the mismatched decoding exponent has a slope of $-1$.

However, it should be taken into consideration that when the block length is $n$, the mismatched decoder is working with a binary code of length $n \cdot L$. The complexity of the maximum-metric decoder is proportional to the number of codewords $2^{n \cdot L \cdot R}$ [8], where $R$ is the rate of the binary code. On the other hand, the number of codewords in the PBICM scheme is $L \cdot 2^{n \cdot R}$ only. In order to assure a fair comparison from the complexity point of view, one has to allow the PBICM scheme to use a block length that is $L$ time the block length of the mismatched decoding scheme. Comparing the error probabilities of both schemes gives $nL\mathbf{E}_r^{\text{PBICM}} = n\mathbf{E}_r^{\text{Mismatched}}$. We therefore define the normalized PBICM error exponent as $L$ times the PBICM error exponent. We conclude that when the complexity is more important (and the latency is less important), the normalized PBICM exponent is the quantity of interest.

It could be claimed, of course, that practical codes used today (such as low-density parity check (LDPC) codes ) will be used and they do not have exponential decoding complexity. On the other hand, such codes do not guarantee an exponentially decaying error probability.

### *B. Comparison with the Mismatched Decoding Exponent*

In the following figures we show the comparison between the PBICM error exponent and the mismatched decoding error exponent [8]. The figures show the (unconstrained) random coding error exponent of the channel, along with the mismatched error exponent and the PBICM random coding error exponent (both normalized and un-normalized).

Figure 10 compares the exponents of 16QAM signaling over the Rayleigh fading channel at SNR = 5dB. Figure 11 shows the same graph, zoomed-in on the capacity region. It can be seen that throughout the entire range of rates between zero and the BICM capacity, the normalized PBICM random coding exponent is higher (better) than the mismatched decoding exponent. Both BICM exponents are above zero for rates below the BICM capacity, and the unconstrained random coding exponent reaches zero at the full channel capacity, as expected. A fact that might be somewhat surprising at first glance is that the normalized PBICM exponent is better than the unconstrained random coding exponent in some rates. While this may seem contradictory, recall that we consider coding schemes with the same maximum-likelihood (or maximum metric) complexity. When normalizing the schemes complexity, PBICM operates with a block length that is $L$ times the block length of the unconstrained scheme, and therefore there is no contradiction. The mismatched decoder never attains higher values than the unconstrained exponent, a fact that is known as the data processing inequality for exponents (see e.g. [8, Proposition 3.2]).

Figure 12 shows a similar picture (zoomed on the capacity in Figure 13). Again, the normalized PBICM outperforms the mismatched decoding exponent for all rates. In this case, the BICM capacity is very close to the full channel capacity, which enables the normalized PBICM to outperform the unconstrained exponent for essentially all rates.

On the Rayleigh fading channel, the same behavior was observed for the range of all practical ranges of SNR for 8PSK, 16QAM and 64QAM signaling: the normalized PBICM exponent outperformed the mismatched decoding exponent.

On the AWGN channel it cannot be claimed that the normalized PBICM exponent outperforms the mismatched exponent, and the other way around is also not true: for 16QAM signaling and a SNR of 0dB (Fig. 14) the normalized PBICM exponent was better, while for a SNR of 5dB the mismatched exponent was better (Fig. 15).

## VI. DISCUSSION

In this paper we have presented *parallel bit-interleaved coded modulation* (PBICM). The scheme is based on a finite-length interleaver and adding binary dither to the binary codewords. The scheme is shown to be equivalent to a binary memoryless channel, therefore the scheme allows easy code design and exact analysis. The scheme was analyzed from an information-theoretical viewpoint, and the capacity, error exponent and the dispersion of the PBICM scheme were calculated.

Another approach for analyzing BICM at finite block length was proposed in [8], where BICM is thought of as a mismatched decoder. Since this BICM setting uses finite length, the random coding error exponent of the scheme can be calculated. In the previous section we have compared the error exponents of PBICM and of the mismatched decoding approach. When the two schemes have the same latency (same block length) the PBICM exponent is inferior to that of the mismatched decoding approach. However, when the complexity of the scheme is considered (or equivalently, when codeword length of the underlying code is the same), PBICM becomes comparable, and generally better over the Rayleigh fading channel.

An important merit of the PBICM scheme is that it allows an easy code design. In PBICM, one has to design a binary code for a memoryless binary channel. In recent years there have developed methods to design very efficient binary codes, such as LDPC codes [10]. When designing LDPC codes, A desired property of a binary channel is that its output will be symmetric. It appears that no matter what channel $W$ we have at hand, the resulting binary channel $\overline{W}$ is always output-symmetric (when the output is the LLR).

Because of its simplicity and easy code design, we conclude that PBICM is an attractive practical communication scheme, which also allows exact theoretical analysis.

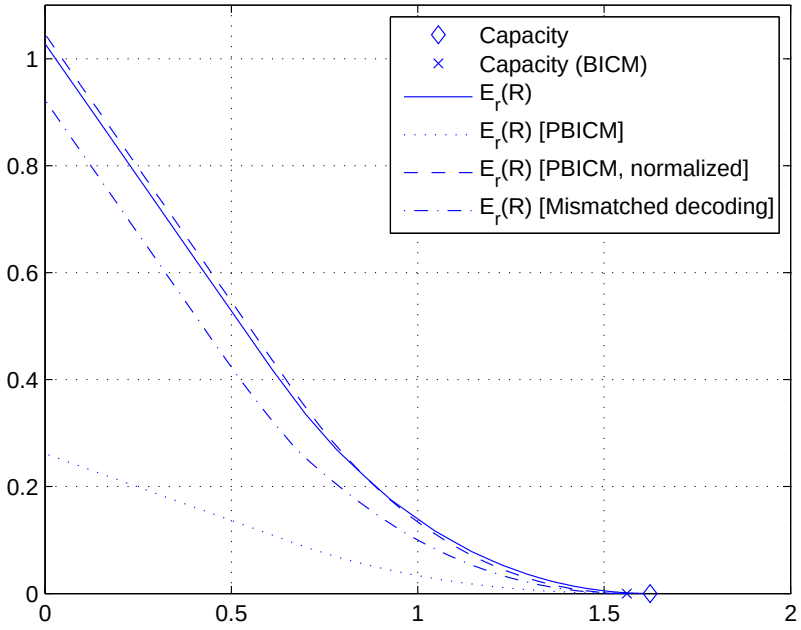


Fig. 10. Random coding exponents over the Rayleigh fading channel with 16-QAM signaling and SNR of 5dB.

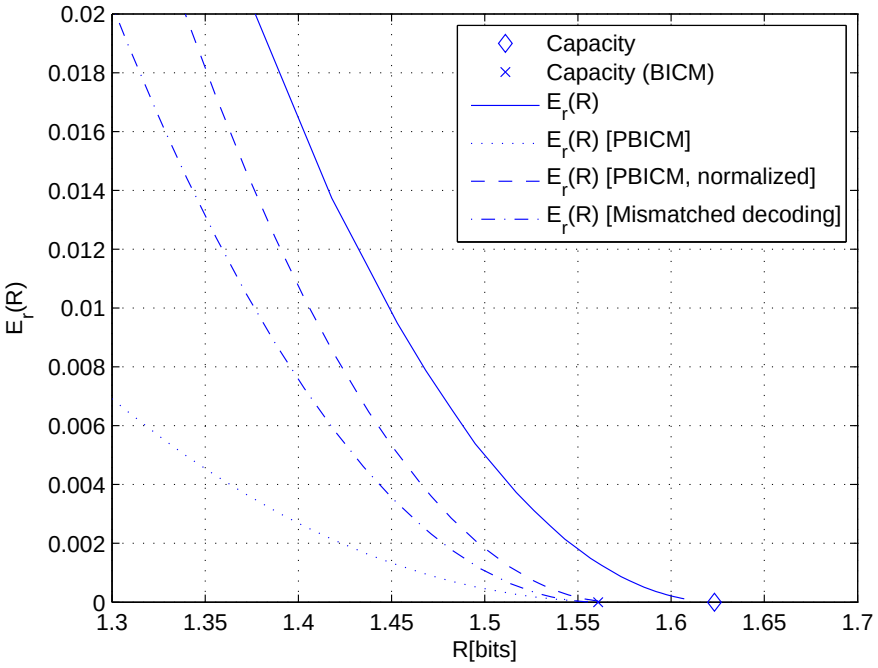


Fig. 11. Random coding exponents over the Rayleigh fading channel with 16-QAM signaling and SNR of 5dB (zoomed on the capacity)

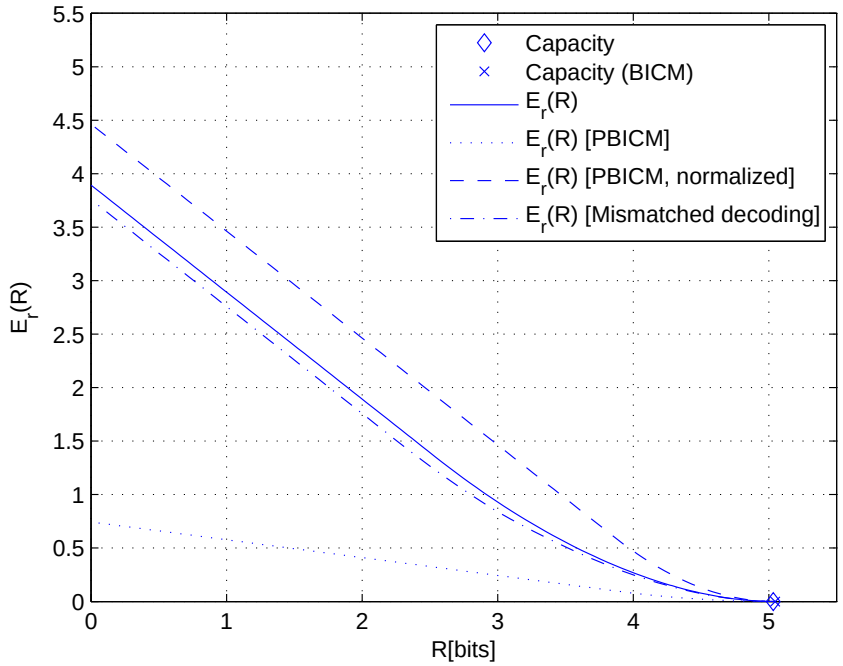


Fig. 12. Random coding exponents over the Rayleigh fading channel with 64-QAM signaling and SNR of 20dB.

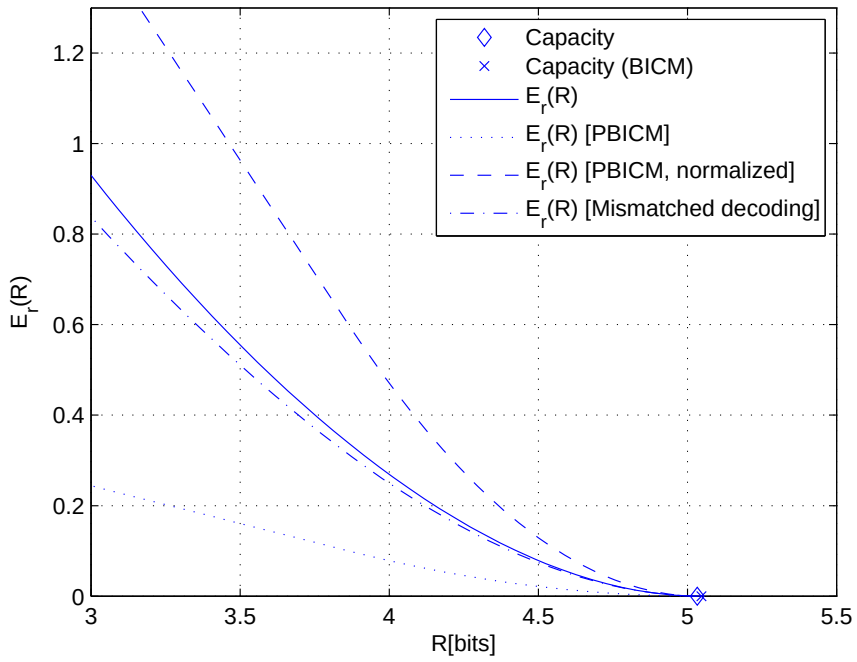


Fig. 13. Random coding exponents over the Rayleigh fading channel with 64-QAM signaling and SNR of 20dB (zoomed on the capacity)

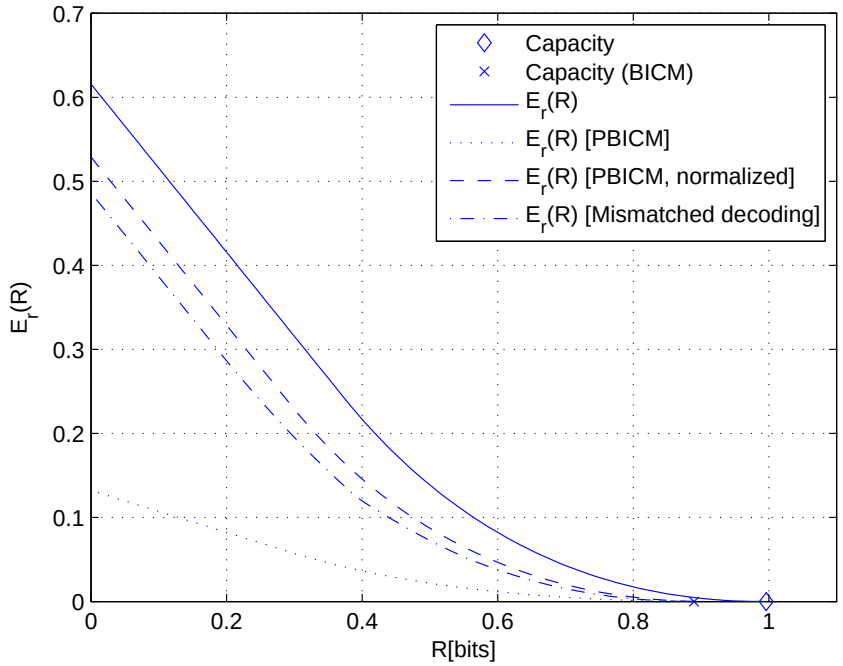


Fig. 14. Random coding exponents over the AWGN channel with 16QAM signaling and SNR of 0dB

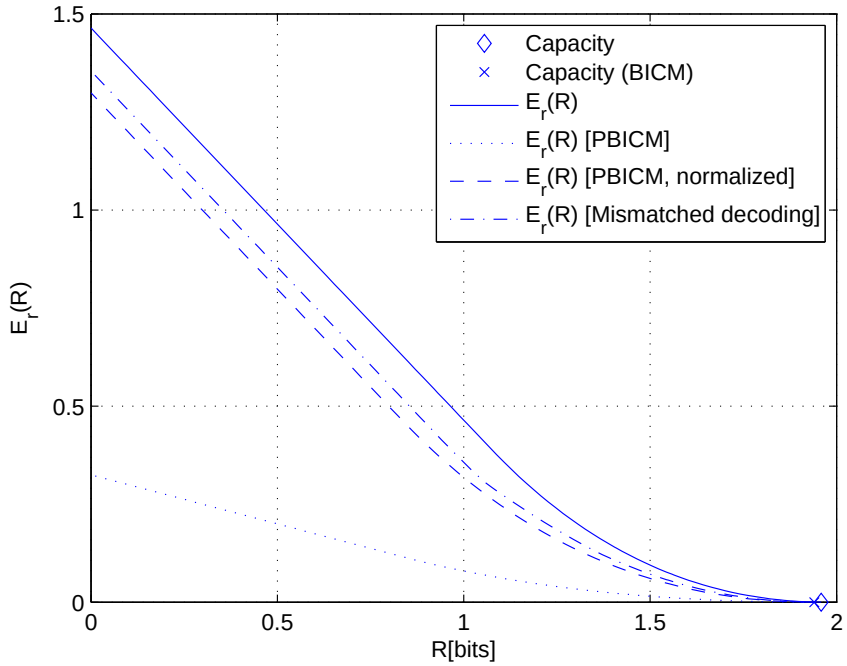


Fig. 15. Random coding exponents over the AWGN channel with 16QAM signaling and SNR of 5dB

Several additional notes can be made:

- The analysis holds for any mapping $\mu$. Finding the mapping that yields the optimal performance at finite lengths is an open question (although Gray mapping is expected to perform well).
- PBICM scheme is composed of, among other things, binary dither. Such tool is used in some cases as a theoretical tool for proving achievability in some problems. In PBICM, it is an essential part of the scheme itself, and even the random capacity proof becomes impossible without it. The main role of the dither is to validate the equivalence of the PBICM scheme to a binary memoryless channel. In addition, the binary dither is the element that symmetrizes the binary channel, which makes the code design easier. This symmetrization property was also noticed by [11] where a similar dither is used with BICM (and termed 'channel adapters'). The code design proposed in [11] rely on the assumption of an ideal interleaver.
- The channel is assumed to be memoryless. This captures many interesting channels, including the AWGN channel, and the memoryless fading channel with and without state known at the receiver (ergodic fading). For slow-fading channels, another interleaver (symbol interleaver) is required in order to transform the slowly fading channel into a fast-fading channel (cf. [2]).

## APPENDIX A
## APPROXIMATION OF THE INVERSE Q-FUNCTION

The following is a useful approximation for the inverse Q-function.

*Lemma 2:*

$$\lim_{\varepsilon\to 0}\left[\frac{(Q^{-1}(\varepsilon))^2}{2\ln\frac{1}{\varepsilon}}\right]=1. \tag{62}$$

*Proof:*

We start with the well known bound on the Q function:

$$\frac{1}{\sqrt{2\pi}x}\left(1+\frac{1}{x^2}\right)e^{-\frac{x^2}{2}}\le Q(x)\le\frac{1}{\sqrt{2\pi}x}e^{-\frac{x^2}{2}} \tag{63}$$

Dividing by the upper bound yields

$$\left(1+\frac{1}{x^2}\right)\le\frac{Q(x)}{\frac{1}{\sqrt{2\pi}x}e^{-\frac{x^2}{2}}}\le 1. \tag{64}$$

Taking the limit $x\to\infty$ gives

$$\lim_{x\to\infty}\frac{Q(x)}{\frac{1}{\sqrt{2\pi}x}e^{-\frac{x^2}{2}}}=1. \tag{65}$$

Since the limit exists, we may take the natural logarithm:

$$\lim_{x\to\infty}\ln\frac{Q(x)}{\frac{1}{\sqrt{2\pi}x}e^{-\frac{x^2}{2}}}=0. \tag{66}$$

$$\lim_{x\to\infty}\ln Q(x)-\ln\frac{1}{\sqrt{2\pi}x}-\ln e^{-\frac{x^2}{2}}=0. \tag{67}$$

Since $\lim_{x\to\infty}\ln Q(x)=-\infty$, we get

$$\lim_{x\to\infty}\frac{\ln Q(x)-\ln\frac{1}{\sqrt{2\pi}x}-\ln e^{-\frac{x^2}{2}}}{\ln Q(x)}=0, \tag{68}$$

which leads to

$$\lim_{x\to\infty}\frac{\ln e^{-\frac{x^2}{2}}}{\ln Q(x)}=\lim_{x\to\infty}\frac{-x^2}{2\ln Q(x)}=1. \tag{69}$$

Since $\lim_{\varepsilon\to 0}Q^{-1}(\varepsilon)=\infty$, we may substitute $x$ with $Q^{-1}(\varepsilon)$, and write

$$\lim_{\varepsilon\to 0}\frac{-(Q^{-1}(\varepsilon))^2}{2\ln\varepsilon}=1, \tag{70}$$

which leads to (62). ■

## APPENDIX B
## BIG-O NOTATION:

As usual, $f(n) = O(\varepsilon_n)$ means that there exist $c > 0$ and $n_0 > 0$ s.t. for all $n > n_0$, $|f(n)| \leq \varepsilon_n$ or equivalently, that

$$-c\varepsilon_n \leq f(n) \leq c\varepsilon_n. \tag{71}$$

$f_n = g_n + O(\varepsilon_n)$ will mean that $f_n - g_n = O(\varepsilon_n)$, which means that $f_n$ can be approximated by $g_n$, up to a factor that is not greater in absolute value than $c \cdot \varepsilon_n$ for some constant $c$.

Sometimes we will be interested in only one of the sides in (71). For that purpose, $f(n) \leq O(\varepsilon_n)$ means that there exist $c > 0$ and $n_0 > 0$ s.t. for all $n > n_0$, $f(n) \leq c \cdot \varepsilon_n$, and $f(n) \geq O(\varepsilon_n)$ will mean that there exist $c > 0$ and $n_0 > 0$ s.t. for all $n > n_0$, $-f(n) \leq c \cdot \varepsilon_n$.

The different combinations of usages of the O notation are listed in the table below.

| Notation | Meaning |
|---|---|
| $f_n = O(\varepsilon_n)$ | $\exists_{c>0,n_0>0}\forall_{n>n_0} \quad \lvert f_n\rvert \leq c \cdot \varepsilon_n$ |
| $f_n = g_n + O(\varepsilon_n)$ | $f_n - g_n = O(\varepsilon_n)$ |
| $f_n \leq O(\varepsilon_n)$ | $\exists_{c>0,n_0>0}\forall_{n>n_0} \quad f_n \leq c \cdot \varepsilon_n$ |
| $f_n \leq g_n + O(\varepsilon_n)$ | $f_n - g_n \leq O(\varepsilon_n)$ |
| $f_n \geq O(\varepsilon_n)$ | $-f_n \leq O(\varepsilon_n)$, or $\exists_{c>0,n_0>0}\forall_{n>n_0} -f_n \leq c \cdot \varepsilon_n$ |
| $f_n \geq g_n + O(\varepsilon_n)$ | $f_n - g_n \geq O(\varepsilon_n)$ |

Note that $f_n \leq O(\varepsilon_n)$ with $f_n \geq O(\varepsilon_n)$ is equivalent to $f_n = O(\varepsilon_n)$, as expected.


## ACKNOWLEDGMENT

Interesting discussions with A. G. i Fàbregas are acknowledged.


## REFERENCES


[1] E. Zehavi, "8-PSK trellis codes for a Rayleigh channel," *IEEE Trans. on Communications*, vol. 40, no. 5, pp. 873–884, May 1992.

[2] G. Caire, G. Taricco, and E. Biglieri, "Bit-interleaved coded modulation," *IEEE Trans. on Information Theory*, vol. 44, no. 3, pp. 927–946, 1998.

[3] R. G. Gallager, *Information Theory and Reliable Communication*. New York, NY, USA: John Wiley & Sons, Inc., 1968.

[4] V. Strassen, "Asymptotische abschätzungen in shannons informationstheorie," *Trans. Third Prague Conf. Information Theory, 1962, Czechoslovak Academy of Sciences*, pp. 689–723.

[5] Y. Polyanskiy, V. Poor, and S. Verdú, "Dispersion of Gaussian channels," in *Proc. IEEE International Symposium on Information Theory*, 2009, pp. 2204–2208.

[6] Y. Polyanskiy, H. Poor, and S. Verdú, "Channel coding rate in the finite blocklength regime," *IEEE Trans. on Information Theory*, vol. 56, no. 5, pp. 2307 –2359, May 2010.

[7] U. Wachsmann, R. F. H. Fischer, and J. B. Huber, "Multilevel codes: Theoretical concepts and practical design rules," *IEEE Trans. on Information Theory*, vol. 45, no. 5, pp. 1361–1391, 1999.

[8] A. Martinez, A. Guillén i Fàbregas, G. Caire, and F. Willems, "Bit-interleaved coded modulation revisited: A mismatched decoding perspective," *IEEE Trans. on Information Theory*, vol. 55, no. 6, pp. 2756–2765, June 2009.

[9] A. Guillén i Fàbregas, A. Martinez, and G. Caire, "Bit-interleaved coded modulation," *Foundations and Trends in Communications and Information Theory*, vol. 5, no. 1-2, pp. 1–153, 2008.

[10] T. J. Richardson, M. A. Shokrollahi, and R. L. Urbanke, "Design of capacity-approaching irregular low-density parity-check codes," *IEEE Trans. on Information Theory*, vol. 47, no. 2, pp. 619–637, 2001.

[11] J. Hou, P. H. Siegel, L. B. Milstein, and H. D. Pfister, "Capacity-approaching bandwidth-efficient coded modulation schemes based on low-density parity-check codes," *IEEE Trans. on Information Theory*, vol. 49, no. 9, pp. 2141–2155, 2003.